DOCUMENT RESUME

ED 317 601 TM 014 696

AUTHOR	Micceri, Theodore
TITLE	Feel No Guilt! Your Statistics Are Probably Robust.
PUB DATE	Apr 90
NOTE	25p.; Paper presented at the Annual Meeting of the American Educational Research Association (Boston, MA, April 16-20, 1990).
PUB TYPE	Reports - Research/Technical (143) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Educational Research; Equations (Mathematics); *Estimation (Mathematics); Mathematical Models; *Psychometrics; *Robustness (Statistics); Statistical Analysis
IDENTIFIERS	Asymmetric Data; Multinomial Models; Parametric Analysis; Research Replication

ABSTRACT

This paper reports an attempt to identify appropriate and robust location estimators for situations that tend to occur among various types of empirical data. Emphasizing robustness across broad unidentifiable ranges of contamination, an attempt was made to replicate, on a somewhat smaller scale, the definitive Princeton Robustness Study of 1972 to determine how closely results produced in a laboratory environment represent the multiple contaminations encountered among real world data. Contaminations included various mixtures of modalities, digit preferences, tail-weights, sample spaces, and asymmetry. Due at least partly to the almost universal presence of asymmetry, the arithmetic mean in particular and L-estimators in general proved comparatively robust for the situations investigated. Most so-called "robust" estimators proved less efficient than the mean even in rather extreme conditions for these multinomial data sets produced by empirical applications of ability and psychometric measures. These findings imply that prior robustness studies that have found the arithmetic mean and its parametric counterparts to be non-robust may be misleading, since the types of theoretical populations investigated in most research studies do not appear to exist among real world psychometric and education data sets. Seven data tables and two graphs are included. (Author/TJH)

Feel No Guilt!

Your Statistics are Probably Robust

Theodore Micceri, Ph.D.

Department of Educational Leadership and
Center for Interactive Technologies, Applications and Research
Assistant Dean's Office
College of Engineering
University of South Florida
Tampa, FL 33620

Paper presented at the AERA annual conference, Boston, MA., April, 1990

ABSTRACT

The current inquiry reports an attempt to identify appropriate and robust location estimators for situations that tend to occur among various types of empirical data. Emphasizing robustness across broad unidentifiable ranges of contamination, an attempt was made to replicate, on a somewhat smaller scale, the definitive Princeton Robustness Study (PRS) of 1972 to determine how closely results produced in a laboratory environment represent the multiple contaminations encountered among real world data. Contaminations included various mixtures of modalities, digit preferences, tail-weights, sample spaces and asymmetry. Thanks at least partly to the almost universal presence of asymmetry, the arithmetic mean in particular and L-estimators in general proved comparatively robust for the situations investigated. Most so-called "robust" estimators proved less efficient than the mean even in rather extreme conditions for these multinomial data sets produced by empirical applications of ability and psychometric measures. These findings imply that prior robustness researches that have found the arithmetic mean and its parametric counterparts to be non-robust may be misleading, since the types of theoretical populations investigated in most researches do not appear to exist among real world Psychometric and Education data sets.

Historically, robustness studies investigating location estimators have found the arithmetic mean to perform poorly. In fact, the seminal Princeton Robustness Study (PRS) of Andrews, Bickel, Hampel, Huber, Rogers and Tukey (1972) claimed the arithmetic mean was the best candidate as the "worst" estimator among 65 investigated. Similar findings are reported in other studies dealing with strictly theoretical data: David and Shu, 1978; Gastwirth and Rubin, 1975; Wegman and Carroll, 1977; Jaeckel, 1971; Ansell, 1973; Carroll, 1979. The implications these studies have for location dependent statistics are somewhat unnerving, especially in view of findings reporting the non-robustness of statistics such as r (Kowalski, 1972; Wainer & Thissen, 1976) and t or F (Blair, 1981; Tan, 1982).

Although none can argue with the findings of the methodologically sound inquiries cited here, the context within which they are conducted raises fundamental questions. Most of these studies have limited themselves to a relatively narrow range of smooth, continuous theoretical distributions. However, there is no reason to expect the parent populations of empirical data to be smooth and continuous. Investigators of real world data consistently find it to differ from theoretical forms and to exhibit discreteness, asymmetry, multimodalities, lumpiness and digit preferences in generally unidentifiable combinations (Allport, 1934; Bradley, 1977; Hill & Dixon, 1982; Micceri, 1989; E.S. Pearson & Please, 1975; K. Pearson, 1895; Stigler, 1977; Tapia & Thompson, 1978; Walberg, Strykowski, Rovai & Hung, 1984). Stigler (1977) points out the weakness in using only simulated mathematical functions to evaluate robustness:

> "... no matter how clever the investigator is in his choice of specifications for sampling distributions, there is no guarantee that the pseudo-samples he generates are actually representative of real data."

Two limited robustness studies involving location estimators have been conducted using real-world data. Stigler (1977), declared the arithmetic mean to be a reasonable estimator and the 10% alpha trimmed mean the optimal statistic for 20 empirical data sets drawn from 18th and 19th century physical science measurements. Hill and Dixon (1982) in a study of four "typical" real world biomedical distributions, recommend the 15% α trimmed mean as a "safe" estimator due to good performance across all the real-world data sets in addition to high relative efficiency at the Gaussian.

The studies cited above dealt with only a few exemplary distributions. The current study attempts to expand the knowledge base of real world robustness and is based on findings from an extensive survey of distributional characteristics for ability and psychometric measures, much of which is reported in Micceri (1989). Twenty-five reputable location estimators were compared in the presence of 37 different distributions. Three pragmatic considerations guided this study's design: (1) the almost universal presence of at least some asymmetry among empirical data, (2) the impact this has on the parameter issue and (3) the concept of robustness across a broad spectrum of distributions because of the complexity found in discrete multinomial distributions. The 37 distributions were chosen to represent ranges of contamination both common and uncommon that have been shown to occur among the psychometric and ability measures that today are so widely used by researchers and decision makers in education, psychology, public health, business and government. Investigators of other types of data suggest that similar contaminations occur elsewhere (Allport, 1934; E.S. Pearson & Please, 1975; K. Pearson, 1895; Stigler, 1977; Tapia & Thompson, 1978; Walberg, Strykowski, Rovai & Hung, 1984). The purpose of this study was to provide a map of performance that may be used by researchers when selecting estimators for empirical investigations. Therefore, the distributions investigated were generally less extreme than

those usually evaluated in robustness research because they were selected to represent the proven and probable, rather than the theoretically possible.

Estimators Included

A group of 25 location estimators including L-estimators, M-estimators, R-estimators and adaptive estimators were compared. L-estimators of location, are linear combination of order statistics. M-estimators of location are solutions, T, of the equation

$$\sum_{j=1}^{n} \psi\left(\frac{x_j - T}{S}\right) = 0$$

where ψ is an odd function and S is estimated from an equation of the form:

$$\sum_{j=1}^{n} \chi\left(\frac{x_j - T}{S}\right) = 0$$

A Huber one-step estimator (P15) as described in PRS was computed. This estimator is specified by a preliminary estimate $\hat{\theta}$ (median), a parameter k (1.5) and a robust scale estimate (the interquartile range/1.35).

Two other Huber M-Estimators (H20, H15) are characterized by a ψ function of the form:

$$\text{(a)} \quad \psi(x;k) = \begin{cases} -k & x < j \\ x & -k < x < k \\ k & k < x \end{cases}$$

and the function

$$\text{(b)} \quad \chi(x) = \psi^2(x;k) - \beta(k)$$

where

$$\beta(k) = \int \psi(x;k)^2 \, \phi(dx)$$

The equations (a) and (b) are solved simultaneously for S and T, iteratively starting with the median and interquartile range/1.35 as initial values.

The Hampel estimators included here are M-estimators using

$$S = \frac{\text{med}\left|x_i - 50\%\right|}{.6745}$$

with ψ given by

$$\psi(x) = \text{sgn } x \cdot \begin{cases} |x| & 0 \le |x| < a \\ a & a \le |x| < b \\ \dfrac{c - |x|}{c - b} \cdot a & b \le |x| < c \\ 0 & |x| \ge c \end{cases}$$

The parameters a, b and c for the three such estimators used here are:

	a	b	c
HMD	2.0	2.0	5.5
22A	2.2	3.7	5.9
12A	1.2	3.5	8.0

A sine fu..ction M-estimate (AMT) suggested by Andrews (Andrews *et al.*, 1972) uses the function:

$$\psi(x) = \begin{cases} \sin\left(\dfrac{x}{2.1}\right) & |x| < 2.1\pi \\ 0 & \text{otherwise} \end{cases}$$

R-estimators are based on ranks, and adaptive estimators respond to characteristics of the sample. Table 1 identifies each estimator and its type. More complete details for 18 of the estimators may be found in the PRS.

Seven estimators not defined in the PRS include four L-estimators: a form of the median for grouped data (P50), the outer-mean (OM) computed as the mean of the trimmings for a 25% α trimmed mean and two Winsorized means (W05 and W15). Originating from the work of C.P. Winsor, Winsorized means identify a proportion of sample points indexed either by n or α. The Winsorized sample points are then assigned the value of that sample point immediately below or

above the last point to be Winsorized, and the arithmetic mean is taken. Additionally, three Hogg-type adaptive estimators Hogg (1974) are based on:

$$Q = \frac{(U(.05)-L(.05))}{(U(.50)-L(.50))}, \quad \text{or} \quad Q_1 = \frac{(U(.20)-L(.20))}{(U(.50)-L(.50))}$$

where $U(\alpha)[L(\alpha)]$ is the mean of the largest (smallest) $[(N+1)\alpha]$ observations. The exact form of the three Hogg adaptive estimators is given below.

$$\text{HG1 } T = \begin{cases} T38 & Q > 3.2 \\ T19 & 2.6 < Q \leq 3.2 \\ M & 2.0 < Q \leq 2.6 \\ OM & Q \leq 2.0 \end{cases} \qquad \text{HG2 } T = \begin{cases} T25 & 1.81 < Q1 \leq 1.87 \\ T25 & 1.81 < Q1 \leq 1.87 \\ T10 & Q1 \leq 1.81 \end{cases}$$

Hogg's adaptive Hampel (Wegman and Carroll, 1977) was defined as:

$$\text{HGH } T = \begin{cases} 22A & Q1 \leq 2.00 \\ 12A & Q1 > 2.00 \end{cases}$$

Performance Measures

To investigate robustness, most researchers evaluate the relative bias and variability of different estimators' sampling distributions. For symmetric distributions, the center of symmetry defines a unique true value (parameter). Under asymmetry, however, no generally agreed upon parameter exists, and the estimand question becomes complex.

Defining a parameter of interest is therefore a central issue in this study. Several authors recommend approaches to this problem. Doksum (1975) suggests using the mean/median interval and the estimand's location within this interval. Ansell (1973) compared estimators against two estimands, the mean and the median of a simulated population. Hill and Dixon (1982) averaged an estimator's variability about the mean and the median. Bickel and Lehmann (1975) recommend the use of that which the estimator estimates in the population (self). Thus, the mean's estimand is μ, and the median's is the population median. Therefore, for the asymmetric data at issue here, at least three estimands need to be considered: μ, median and self.

Under asymmetry, μ is pulled in the direction of greatest skew, while the median remains central. A statistician working with an empirical data set is most likely interested in an estimate within this mean/median interval (Doksum, 1975; Bickel & Lehmann, 1975). Most robust estimates are more likely to estimate the median than μ in this situation, and in fact probably estimate a parameter between the two.

TABLE 1

Selected Location Estimators and Their Identifying Labels

Label	Estimator
L-Estimators	
M	Arithmetic Mean
T05	5% Alpha Trimmed Mean
T10	10% Alpha Trimmed Mean
T15	15% Alpha Trimmed Mean
T25	25% Alpha Trimmed Mean
MED	Median Continuous (ungrouped) Data
P50	Median Categorical (grouped) Data
W05	5% Winsorized Mean
W15	15% Winsorized Mean
GAS	Gastwirth's Trimean
TRI	Trimean
CST	Iteratively ' -skipped trimean
OM	Outmean (25%)
M-Estimators	
P15	Huber One-step, k=1.5, start=median
H20	Huber proposal 2, k=2.0
H15	Huber Proposal 2, k=1.5
HMD	Hampel M, PSI bends at 2.0,2.0,5.5
22A	Hampel M, PSI bends at 2.2,3.7,5.9
12A	Hampel M, PSI bends at 1.2,3.5,8.0
AMT	Andrew's sin function M-estimate
R-Estimator	
HL	Hodges-Lehmann R-estimator
Adaptive Estimators	
JOH	John's adaptive estimator
HG1	Hogg's adaptive trimmed mean based on Q
HG2	Hogg's adaptive trimmed mean based on Q1
HGH	Hogg's adaptive Hampel M, based on Q1

For this reason an additional estimand was computed to represent this desirable parameter located somewhere between μ and the median under asymmetry. Termed E-robust (estimate of the robust), it was computed by taking the 15% α trimmed mean of all estimands. The expected value for this statistic under asymmetry is somewhere between μ and the median, a closer approximation to that value many robust estimators seek in the population, and perhaps a value of most interest to many researches. This approach of evaluating an estimator's performance against four different estimands provides information of great use to the practitioner who must decide which point within the mean/median interval is of most interest in a specific research.

In the following discussions, bias is defined as the distance between the mean of a sampling distribution and the value of the distribution's (hereafter termed pseudo-population) target parameter θ. Variability is the mean square error of a sampling distribution, an absolute measure of dispersion, and MSE is the mean square error of a sampling distribution about the pseudo-population target parameter θ.

Distributions

The distributions investigated by Micceri (1989) were exclusively multinomial in nature and were frequently characterized by multiple forms of contamination including various multimodalities and digit preferences as well as diverse tail weights and degrees of asymmetry. Figure 1 depicts two common examples of psychometric distributions that exhibit mixed asymmetric contamination combined with ceiling and floor effects.

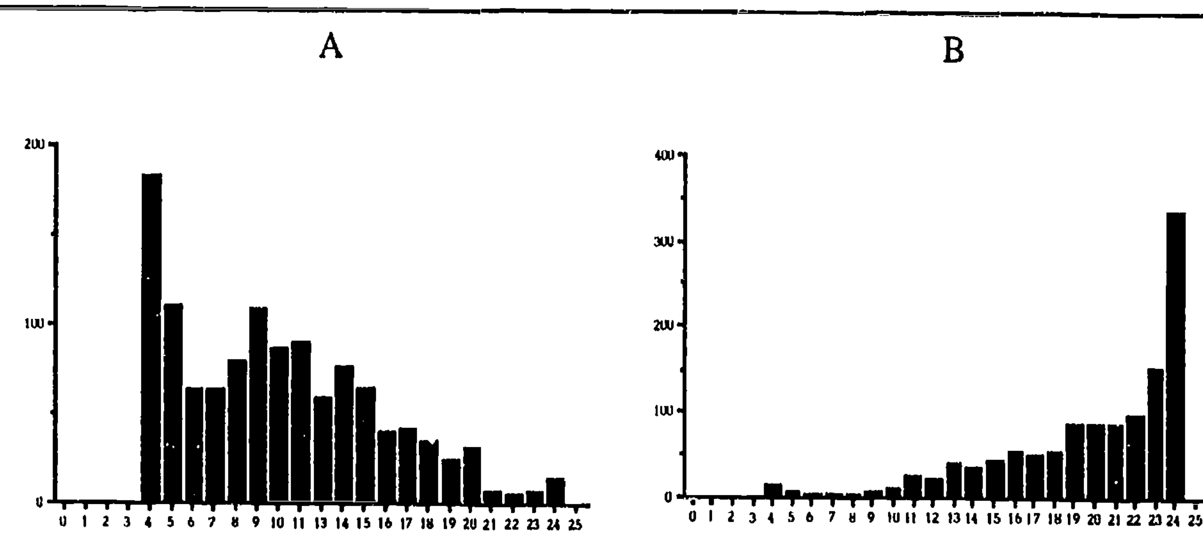

Figure 1: Two asymmetric empirical distributions: Panel A (N=1258), Panel B (N=1231)

Micceri (1989) used robust estimates of tail weight and asymmetry to classify a broadly representative group of large sample empirical distributions into one of four categories defining symmetry or the lack thereof, and one of six categories defining various levels of tail weight (Table 2). The author notes (p. 158):

> Cut points were set arbitrarily, and those defining "moderate contamination" of either tail weight or asymmetry were selected only to identify distributions as DEFINITELY non-Gaussian. The moderate contamination cut points (both symmetric and asymmetric) were set at 5% and 15% contamination based on the support for the alpha trimmed mean and trimmed t in the research literature.

A sample of 35 distributions was originally selected to represent the varying combinations of asymmetry, tail weight, modality and lumpiness found among the real life pseudo-populations represented in Table 2. To assure that the distributions well represented their parent populations, either 1000 cases or at least 10 cases for each possible sample point was required for inclusion in this study. Pseudo-population sample sizes ranged from 400 to 4994, with 28 (75%) including more than 1000 cases. Sample spaces ranged from 4 to 262.

In an attempt to thoroughly investigate traditional issues, both symmetric and heavy tailed symmetric distributions were oversampled relative to the proportions found among real data by Micceri (1989). First run findings for the relatively symmetric, extremely heavy tailed situations contradicted prior researches. Therefore, two additional such pseudo-populations were drawn as a cross validation.

Real world distributions present interpretation problems due to their multiple contaminations. Because of this and the complexity created by 37 different situations, these pseudo-populations were not treated as unique entities, but rather were considered to represent different mixtures of contamination: sample space, tail weight, asymmetry, multimodality and/or digit preference. For example, each of the four different pseudo-populations representing moderate contamination of tail weight and asymmetry had a different combination of sample space, modality and digit preference. One was unimodal and smooth, one multimodal and smooth, one unimodal and lumpy, and one multimodal and lumpy, with sample spaces respectively of 21, 31, 45 and 137. Each sampled cell of Table 2 that includes four pseudo-populations contains a similar mixture of contaminations. Thus, a variety of factors was averaged within a specific combination of tail weight and symmetry conditions to assess general performance in the presence of real world complexities.

Results for variability, bias and MSE were computed separately for each pseudo-population. These were then combined within five major contamination strata each containing between two and four substrata:

(1) four relatively symmetric tail weight groups: lighter than Gaussian, about Gaussian, moderately heavy tailed and extremely heavy tailed (cross validated once);

(2) four levels of symmetry: relatively symmetric, moderately asymmetric, extremely asymmetric and exponentially asymmetric;

(3) three categories comparing increasingly greater mixtures of asymmetric contamination and tail weight (Gaussian level, moderate contamination of both tail weight and asymmetry, and extreme contamination of both tail weight and asymmetry);

(4) three different types of symmetric modality characteristics: smooth unimodal, lumpy unimodal and smooth multimodal and

(5) symmetric and asymmetric small sample space situations (fewer than 10 scale points).

The purpose of creating so many categories was to provide a matrix of results from which specific applications can be evaluated. The reason for collapsing complex combinations within cells was twofold: (1) to simplify interpretation and reporting and (2) to more closely represent the indefinable complexity of the real world, where isolated contaminations are rarely found.

One thousand estimates were computed for each of the 25 location estimators for each of the 37 pseudo-populations for samples of size 5, 10, 20 and 40 using the IMSL uniform random number generator (GGUD). Algorithms for location estimators came primarily from the PRS, with minor modifications (mainly in transferring from FORTRAN 66 to FORTRAN 77). More elegant algorithms for three estimators (CST, P15 and OM) result from the work of Keller-McNulty & Higgins (1987). Subroutines for P50 were written by the author.

Since sample spaces, means, medians and variability differed for each pseudo-population; variability, bias and MSE were standardized by the respective population's standard deviation. Following the tradition established by the PRS, deficiencies were then used to compare estimators on bias, variability and MSE. Deficiencies are defined as 1 - efficiency, where:

$$\text{efficiency} = \frac{\text{performance of optimum estimator}}{\text{performance of estimator of interest}} .$$

Although not symmetric about zero (Wegman and Carroll, 1977), deficiencies show deviations from optimum, emphasize values close to 0.00, are easily understood and cause only limited problems with accuracy, at least when below .25.

Table 2

Values of Tail Weight and Asymmetry For All Pseudo-Populations

Values of Tail Weight	Values of Asymmetry: Near Symmetry n		Values of Asymmetry: 5% Asym. Contam. n		Values of Asymmetry: 15% Asym. Contam. n		Values of Asymmetry: Exponential n		Totals N		Pct.
Uniform	0	(0)*	4	(0)	5	(1)	5	(0)	14	(1)	3.2%
Less than Gaussian	21	(2)	33	(1)	8	(1)	3	(0)	65	(4)	14.8%
Near Gaussian	30	(6)	29	(2)	7	(3)	1	(1)	67	(12)	15.2%
5% Symmetric Contam 2 std dev from mean	30	(2)	35	(4)	11	(1)	2	(1)	78	(8)	17.8%
15% Symmetric Contam. 3 std dev from mean	41	(4)	64	(1)	35	(1)	3	(0)	143	(5)	32.5%
Double Exponential	3	(2)	14	(1)	20	(2)	36	(1)	73	(5)	16.6%
Totals	125	(14)	179	(9)	86	(9)	50	(3)	440	(35)	
Pct.	28.4%		40.7%		19.6%		11.4%				100.0%

* Number of pseudo-populations included in comparison of location estimators

Results and Discussion

Much like other studies investigating asymmetric situations (i.e. David and Shu, 1978) the results of this study for variability and bias tended to contradict each other. In relatively symmetric situations, the less variable estimators for all four parameters tended to exhibit greater bias, and *vice versa.* Both absolute standardized variability and differences among estimators on variability tended to be considerably greater than either absolute standardized bias or differences among estimators on bias. Variability of the sampling distributions averaged respectively for samples of size: 5 (.50s_X of the pseudo-population), 10 (.35s_X), 20 (.25s_X) and 40 (.20s_X). Although OM exhibited rather extreme bias from the median in the exponential situation, bias from E-robust was generally small (> .05s_X).

Overall, M and OM were the least variable estimators. For almost every situation, M was either the least variable or one of the least variable estimators. All estimators were both relatively and absolutely quite unbiased relative to their own estimand (self) for all sample sizes and distribution shapes. As asymmetry increased, M exhibited increasingly greater bias from both the median and E-Robust, yet generally remained a comparatively unbiased estimator of μ. As asymmetry increased, most "robust" estimators exhibited less bias from the median than μ and in fact appeared to estimate E-Robust. For the small sample situation (n=5), under both symmetric and asymmetric conditions, several rather drastic estimators (MED, P50, JOH, GAS, TRI, CST, 12A and AMT) performed well in terms of bias (particularly from the median), but exhibited relatively great variability. As sample size increased, these estimators tended to lose their advantage of lesser bias while retaining their disadvantage of greater variability. For almost every situation, little change in relative performance occurred past sample size 20. Detailed results for bias and variability can be found in Micceri (1987).

RESULTS FOR MSE

Because results for bias and variability tended to produce different conclusions and since most estimators proved relatively unbiased estimates of self, further emphasis here is placed on the MSE of each estimator's sampling distribution about three population parameters: μ, median and the "in between" E-Robust. Table 3 shows the **most** efficient estimators for each sampling situation only for samples of size 5 and 20 since results were quite consistent as sample size increased. Only eight estimators proved most efficient in more than a single situation (M, OM, T05, T10, T15, T25, H15 and H20). M (58 of 96 situations) is most frequently the most efficient estimator. Under symmetry, M is optimum for all parameters for every tail weight situation except the light tailed and moderately heavy tailed situations. As estimators of μ, one of M, OM or T05 is always most efficient. As estimators of E-Robust, M(19), OM(6), T05(3), T10(2) and T15(2) dominate. As estimators of the median, M, OM and T05 are most efficient for the majority of symmetric situations, with T15 and T25 proving most efficient for extreme asymmetry or combinations of extreme asymmetry and heavy tail weight.

Where they are more efficient, traditional robust estimators become increasingly more efficient as sample size increases. However, almost all statistics performed fairly well in an absolute sense for the large sample situation. Tables 4 to 6 show the deficiencies of the seven overall best estimators (OM, M, T05, T10, T15, H20 and H15) from the optimum estimator of the three parameters across the 16 sampling situations for samples of size 5 and 20.

In the following discussion, estimators in parenthesis are optimum for the situation. Table 4 shows that as an estimator of μ, M is either most efficient or within 1% of the most efficient estimator for all situations except the symmetric light tailed (OM), and the symmetric small sample space (OM). Table 5 shows that as an estimator of E-Robust, M exhibits deficiencies greater than 1% only for symmetric light tails (OM), small sample spaces both symmetric (OM, H15) and asymmetric (T05) and exponential asymmetry or extreme mixed contamination (T15, T10). As an estimator of the median (Table 6), results are similar with greater deficiencies, particularly as either sample size or asymmetric contamination increases.

As expected, Tables 4 to 6 show that for all relatively symmetric situations all estimators attempt to estimate approximately the same parameter. Under relative symmetry, only for light tailed or small sample space pseudo-populations is M more than 1% deficient from the optimum estimator.

Results under asymmetry are quite consistent. As contamination increases, M remains the most efficient estimator of μ, however, as an estimator of E-Robust or the median, M's deficiency increases as asymmetry increases. Among trimmed means, as asymmetry and/or sample size increase, the more efficient estimators tend to be those that trim increasingly greater percentages of the sample. Note that the mixed contamination class includes distributions exhibiting both extreme and exponential asymmetry, thus having on average less extreme asymmetry than those classified as exponentially asymmetric.

TABLE 3

Most Efficient Estimator of Three Pseudo-Population Parameters for Samples of Size 5 and 20 Across a Variety of Sampling Situations

		Most Efficient Estimator					
		Sample Size 5			Sample Size 20		
Pseudo-Population Type	N	Abt. Mean	Abt. E-Rb	Abt. Med	Abt. Mean	Abt. E-RB	Abt. Med
SYMMETRIC							
Light Tail	2	OM	OM	OM	OM	OM	OM
Gaussian	6	M*	M	M	M	M	M
Mod. Tails	2	T05	T05	T05	T05	T05	T05
Heavy Tails	6	M	M	M	M	M	M
SYMMETRY TO ASYMMETRY							
Symmetric	14	M	M	M	M	M	M
Mod. Asym.	9	M	M	M	M	M	H20
Ext. Asym.	9	M	M	M	M	W05	T15
Expn. Asym.	3	T05	T15	T25,GAS	M	T10	T25
MIXED CONTAM.†							
Gaussian	6	M	M	M	M	M	M
Mod. Contam.	4	M	M	M	M	M	M
Ext. Contam.	4	T05	T10	T15	M	T15	T25
MODALITY AND DIGIT PREF. UNDER SYMMETRY							
Smth. Unimode	5	OM	M,OM	OM	OM	M	M
Lmpy. Unimode	4	M	M	M	M	M	M
Smth. Mltimod	2	M	M	M	OM	M	M
FEWER THAN 10 SCALE POINTS							
Symmetric	3	OM	OM	OM	OM	OM	OM
Asymmetric	3	M	M	H15	M	T05	T15

* At sample size 5, H20 and M are equivalent.

† Mixed contamination of both tail weight and asymmetry.

Table 4

Deficiencies of Seven Estimators from the
Most Efficient Estimator of the Arithmetic Mean for Sixteen Situations

Pseudo-Populations		Percent Deficiency of MSE from Optimum Estimator													
		Sample Size 5							Sample Size 5						
Type	N	OM	M	T05	T10	T15	H20*	H15	OM	M	T05	T10	T15	H20	H15
SYMMETRIC															
Light Tail	2	0	16	19	23	28	16	22	0	27	36	43	46	27	33
Gaussian	6	2	0	1	3	6	0	4	1	0	3	6	9	1	4
Mod. Tails	2	9	1	0	1	1	1	2	12	1	0	6	1	1	1
Heavy Tails	6	9	0	1	1	3	0	2	8	0	2	4	5	1	3
SYMMETRY TO ASYMMETRY															
Symmetric	14	2	0	1	3	5	0	3	2	0	3	6	8	1	3
Mod. Asym.	9	3	0	1	2	5	0	3	5	0	3	6	9	1	3
Ext. Asym.	9	0	0	2	4	8	0	4	11	0	6	13	18	4	9
Expn. Asym.	3	19	1	0	1	4	1	2	39	0	6	16	25	5	12
MIXED CONTAM.† TAIL WT & ASYM															
Gaussian	6	2	0	1	3	6	0	4	1	0	3	6	9	1	4
Mod. Contam.	4	2	0	1	3	5	0	4	3	0	3	7	10	1	4
Ext. Contam.	4	13	1	0	1	3	1	3	31	0	6	13	19	6	12
MODALITY AND DIGIT PREFERENCE UNDER SYMMETRY															
Smth. Unimode	5	0	1	1	3	7	1	4	0	1	4	8	11	1	4
Lmpy. Unimode	4	7	0	1	1	3	0	2	9	0	1	3	4	1	2
Smth. Mltimod	2	1	0	1	3	6	0	3	0	1	4	8	11	2	5
FEWER THAN 10 SCALE POINTS															
Symmetric	3	0	5	7	10	15	5	9	0	8	14	20	24	9	15
Asymmetric	3	1	0	3	5	11	0	0	16	0	8	18	26	5	11

* At sample size 5, H20 and M are equivalent.
† Mixed contamination of both tail weight and asymmetry.

Table 5

Deficiencies of Seven Estimators from the
Most Efficient Estimator of E-Robust for Sixteen Situations

Pseudo-Populations		Percent Deficiency of MSE from Optimum Estimator													
		Sample Size 5							Sample Size 20						
Type	N	OM	M	T05	T10	T15	H20*	H15	OM	M	T05	T10	T15	H20	H15
SYMMETRIC															
Light Tail	2	0	14	17	20	25	14	19	0	20	27	33	37	21	25
Gaussian	6	2	0	1	3	6	0	4	1	0	3	6	9	1	4
Mod. Tails	2	9	1	0	1	2	1	2	12	1	0	1	2	1	1
Heavy Tails	6	6	0	1	1	3	0	2	9	0	2	3	5	1	3
SYMMETRY TO ASYMMETRY															
Symmetric	14	2	0	1	3	5	0	3	3	0	3	6	8	1	3
Mod. Asym.	9	4	0	1	2	4	0	3	9	0	1	4	6	1	2
Ext. Asym.	9	3	0	1	3	6	0	3	21	1	1	3	7	1	2
Expn. Asym.	3	27	5	3	1	0	5	1	56	22	6	0	1	14	6
MIXED CONTAM.† TAIL WT & ASYM															
Gaussian	6	2	0	1	3	6	0	4	1	0	3	6	8	1	4
Mod. Contam.	4	3	0	1	2	5	0	3	6	0	2	5	7	1	4
Ext. Contam.	4	18	2	1	0	1	2	1	44	12	4	1	0	5	3
MODALITY AND DIGIT PREFERENCE UNDER SYMMETRY															
Smth. Unimode	5	0	0	1	3	7	0	4	1	0	4	8	11	1	4
Lmpy. Unimode	4	7	0	1	1	3	0	2	10	0	1	2	4	1	2
Smth. Mltimod	2	1	0	1	3	6	0	3	1	0	4	7	10	2	5
FEWER THAN 10 SCALE POINTS															
Symmetric	3	0	5	7	10	14	5	8	0	6	12	17	21	7	11
Asymmetric	3	7	0	3	4	8	1	1	30	3	0	5	12	2	2

* At sample size 5, H20 and M are equivalent.
† Mixed contamination of both tail weight and asymmetry.

Table 6

Deficiencies of Seven Estimators from the
Most Efficient Estimator of the Median for Sixteen Situations

Pseudo-Populations		Percent Deficiency of MSE from Optimum Estimator														
		Sample Size 5							Sample Size 20							
Type	N	OM	M	T05	T10	T15	H20*	H15	OM	M	T05	T10	T15	H20	H15
SYMMETRIC																
Light Tail	2	0	14	16	20	25	14	19	0	19	26	32	36	20	24
Gaussian	6	2	0	1	3	6	0	4	1	0	3	6	9	1	4
Mod. Tails	2	9	1	0	1	2	1	2	12	1	0	1	2	1	1
Heavy Tails	6	6	0	1	1	3	0	2	9	0	2	3	4	1	2
SYMMETRY TO ASYMMETRY																
Symmetric	14	2	0	1	3	5	0	3	3	0	3	6	8	1	3
Mod. Asym.	9	5	0	1	2	4	0	3	12	1	1	2	4	0	1
Ext. Asym.	9	7	0	1	1	3	0	1	30	8	3	1	0	6	3
Expn. Asym.	3	33	12	9	6	2	12	6	61	37	23	12	4	30	21
MIXED CONTAM.† TAIL WT & ASYM																
Gaussian	6	2	0	1	3	6	0	4	1	0	3	6	9	1	4
Mod. Contam.	4	3	0	1	2	5	0	3	7	0	2	4	6	1	2
Ext. Contam.	4	21	5	3	1	0	5	2	51	26	17	10	5	22	16
MODALITY AND DIGIT PREFERENCE UNDER SYMMETRY																
Smth. Unimode	5	0	1	1	3	7	1	3	0	0	4	8	10	1	4
Lmpy. Unimode	4	7	0	1	1	3	0	2	2	0	1	2	3	1	2
Smth. Mltimod	2	1	0	1	3	6	1	3	1	0	4	7	10	2	5
FEWER THAN 10 SCALE POINTS																
Symmetric	3	0	5	7	10	14	5	8	0	6	11	16	20	7	11
Asymmetric	3	13	3	3	2	3	3	0	41	16	7	2	0	12	8

* At sample size 5, H20 and M are equivalent.
† Mixed contamination of both tail weight and asymmetry.

OM exhibits relatively low deficiencies until contamination becomes extreme. H20 performs much like M, with slightly lower deficiencies for the extremely contaminated situations and greater deficiencies for all other situations. T05 exhibits slightly greater deficiencies than M for those situations in which OM is optimal, and somewhat lower deficiencies when estimating either the median or E-Robust under extreme contamination. T10 and T15 exhibit comparatively poor properties for the light tailed symmetric situations and are almost never comparatively efficient estimators of the mean. However, as estimators of E-Robust or the median (particularly T15) under all non-symmetric situations they perform well. H15 exhibits at least small deficiencies (3% to 5%) almost everywhere, and appears to be a slightly less efficient estimator of E-Robust than T05 under most conditions.

Limited scale point situations were not investigated thoroughly; however, among the three relatively symmetric, light tailed pseudo-populations, M exhibits noticeable deficiencies from OM in the estimation of all pseudo-parameters. Even in the asymmetric small sample space situations, OM is an efficient estimator of μ for sample size 5, and a reasonably efficient estimator of E-robust. These findings support expectations (Wegman and Carroll, 1977). H15 and T15 are optimum estimators for the median in the asymmetric, small sample space situation respectively for samples of size 5 and 20.

Table 7 shows the deficiencies of the previously tabled estimators under two extremely asymmetric situations. It is clear from this table that as an estimator of μ, M stands out for both of these situations, and as an estimator of both E-Robust and the median, T15 stands out. Thus, for the researcher expecting an extremely asymmetric distribution and wishing to estimate the most commonly occurring score, T15 appears the most appropriate estimator. Micceri (1989), found that 52% of the psychometric distributions investigated to exhibit this level of asymmetry (p. 163).

TABLE 7

Deficiencies from Most Efficient Estimators for Two Forms of Extreme Asymmetric Contamination for Samples of Size 20

	Extreme Asymmetry and Tail Weight			Exponential Asymmetry		
	Mean	E-Rb	Med	Mean	E-Rb	Med
M	0	12	26	0	22	37
OM	31	44	51	39	56	61
T05	6	4	17	6	6	23
T10	13	1	10	16	0	12
T15	19	0	5	25	1	4
H20	6	5	22	5	14	30
H15	12	3	16	12	6	[illegible]

Conclusions and Implications

It is obvious that M and T05 are the overall most efficient estimators. M is an excellent estimator of the population mean except in the relatively symmetric light tailed situation, where OM dominates. Additionally, T15 is a good choice to estimate either the median or E-Robust under extreme asymmetry, at least for the pseudo-populations studied, which should be a fairly representative sample of distribution types found among ability and psychometric measures. For samples of size 5, the rather radical estimators (MED, P50, JOH, GAS, TRI, CST, 12A, AMT) perform well in terms of bias, but exhibit severe variability and therefore do not appear good choices to estimate any parameter investigated. Those estimators that perform worst overall, again, largely due to their variability, are MED, JOH, GAS, TRI, CST. The L-estimators outperform the M-estimators, R-estimator and adaptive estimators, with few exceptions. All estimators usually exhibited small absolute biases from their personal estimands (self).

Although the results of this study appear to differ from prior researches, the differences occur mostly because of approach and context. Most prior researches sought elegant solutions. Note that Andrews *et. al.* (1972, p. 109) investigated two asymmetric situations and found the arithmetic mean to be the least variable (best) estimator for both. This finding was never again mentioned because the authors "... were not able to agree, either between or within individuals, as to the criteria to be used." (p. 226) Similar findings are reported by David & Shu (1978). Hill and Dixon (1982) averaged MSE about both μ and the median. The arithmetic mean proved the most efficient estimator for their single relatively symmetric distribution, but when variability was averaged about both m and the median for three exceedingly asymmetric populations, the arithmetic mean proved less efficient than the trimmed means and others. This probably resulted because the mean was an efficient estimator of μ and a relatively inefficient estimator of the median, while the robust estimators were moderately inefficient estimators of both parameters. This penchant for seeking elegant solutions can prove misleading in the decidedly inelegant world of reality. Although more complicated and time consuming, treating each of three possible parameters that span the mean/median interval independently provides a clearer perspective on an estimator's performance in specific situations. From a practitioners standpoint, a good estimate of one parameter or two good estimates of two different parameters (μ and median) are probably more useful than a single estimates that combines error about two different parameters.

Another very interesting finding involves symmetric distributions. Results for extremely heavy tailed situations, in which M proved the most efficient estimator, differed substantially from historic robustness studies (particularly Andrews *et. al.*, 1972), where M tended to be one of the least efficient estimators. For this reason, detailed investigations of the relevant pseudo-populations' distributional characteristics were conducted, and brought to light the fact that the two "moderately" heavy tailed examples were among the 10 most symmetric of the 440 distributions studied in Micceri (1989). Their tails were almost perfectly balanced even at the 90th, 95th and

97th percentiles. Among the 44 relatively symmetric extremely, heavy tailed distributions under study, only one exhibited such perfect symmetry. Prior robustness researches investigating have dealt exclusively with perfectly symmetric, theoretical populations. In the universally asymmetric real world, results differ. This explains why the moderately heavy tailed cases produced results more in keeping with earlier studies than did the extremely heavy tailed pseudo-populations, in that M was not the most efficient estimator. Under perfect symmetry and increasingly heavy tail weights, M becomes increasingly less robust. However, in the real world, as tail weight increases, it tends to do so in a nonsymmetric fashion Micceri (1989).

In addition, a general lack of distant extremes occurred among the relatively symmetric, heavy tailed distributions, which rarely had remote cases exceeding four standard deviations from the pseudo-population median. This further supports the notion that extremely long-tailed, symmetric distributions are quite rare in the real world (Micceri, 1989; Walberg *et al.*, 1984). Certainly the radically asymmetric distributions explored here included distant extremes, with several of the pseudo-populations having remote observations more than six standard deviations from the median.

Studies such as the PRS have investigated robust estimators in theoretical realms, and particularly in the face of types of contamination for which those estimators were designed. Almost none of the empirical distributions studied evidence either smoothness or continuity. Certainly characteristics similar to the Cauchy, rectangular and T with one degree of freedom were almost nonexistent. Despite claims for their robustness at Normality, few estimators that performed well in Andrews, *et al.* (1972) and other researches also performed well for the empirical, multiply contaminated populations considered here. If is quite possible that these results will generalize to tests of significance and regression. Some preliminary investigations by the author and a colleague suggest that even Pearson's r proves quite robust for these real world types of contaminations. These findings suggest that much robustness research has been conducted on smooth and rather extreme theoretical distributions that simply fail to exist in the real world. Among the more moderate and complex contaminations that occur in distributions of Education and Psychology, it is probable that typically applied test statistics are relatively robust, and that we only rarely need feel guilty for our use of them.

References

Allport, F. M. (1934). The J-curve hypothesis of conforming behavior. Journal of Social Psychology, 5, 141-183.

Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J., Rogers, W. H., & Tukey, J. W. (1972). Robust estimates of location survey and advances. Princeton, N.J.: Princeton University Press.

Ansell, M. J.G. (1973). Robustness of location estimators to asymmetry. Applied Statistics, 22, 249-254.

Bickel, P J. and Lehmann, E. L. (1975). Descriptive statistics for nonparametric models II. Loca ion. The Annals of Statistics, 3, 1045-1069.

Blair, R. C. (1981). A reaction to "Consequences of failure to meet assumptions underlying the fixed effects analysis of variance and covariance." Review of Educational Research, 51, 499-507.

Bradley, J. W. (1977). A common situation conducive to bizarre distribution shapes. The American Statistician, 31, 147-150.

Carroll, R.J. (1979). On estimating variances of robust estimators when the errors are asymmetric. Journal of the American Statistical Association, 74,674-679.

David, H. A. and Shu, V. S. (1978). Robustness of location estimators in the presence of an outlier. In H. A. David (Ed.) Contributions to Survey Sampling and Applied Statistics, New York: Academic Press, 235-250.

Doksum, K. A. (1975). Measures of location and asymmetry. Scandinavian Journal of Statistics, 2, 11-22.

Gastwirth, J. L. and Rubin, H. (1975). The behavior of robust estimators on dependent data. The Annals of Statistics, 3, 1070-1100.

Hill, M. and Dixon, W. J. (1982). Robustness in real life: A study of clinical laboratory data. Biometrics, 38, 377-396.

Hogg, R. L. (1974). Adaptive robust procedures: A partial review and some suggestions for future applications and theory. Journal of the American Statistical Association, 69, 909-927.

Jaeckel, L. A. (1971). Robust estimates of location: Symmetry and asymmetric contamination. The Annals of Mathematical Statistics, 42, 1020-1034.

Keller-McNulty, S. & Higgins, J. J. (1987). The effect of tail weight and outliers on power and type I errors of robust permutation tests. Communications in Statistics, B16, 17-36.

Kimber, A. C. (1982). Tests for many outliers in an exponential sample. Applied Statistics, 31, 263-271.

Kowalski, C. J. (1972). On the effects of non-normality on the distribution of the sample product-moment correlation coefficient. Applied Statistics, 21, 1-12.

Micceri, T. (1987). Testing for normality and evaluating the relative robustness of location estimators for empirical distributions derived from achievement tests and psychometric measures. Unpublished doctoral dissertation, University of South Florida

Micceri, T. (1989). The unicorn, the normal curve and other improbable creatures. Psychological Bulletin, 105:1, 156-166.

Pearson, E.S. & Please, N.W. (1975). Relation Between the Shape of Populations Distributions and the Robustness of Four Simple Test Statistics. Biometrika, 62, 223-241.

Pearson, K. (1895). Contributions to the mathematical theory of evolution: II. Skew variation in homogeneous material. Philosophical Transactions of the Royal Society, Ser. A, 186, 343-414.

Stigler, S. M. (1977). Do robust estimators work with real data?. The Annals of Statistics, 5, 1055-1098.

Tan, W. Y. (1982). Sampling distributions and robustness of *t*, *F* and variance-ratio in two samples and ANOVA models with respect to departure from normality. Communications in Statistics, A11, 2485-2511.

Tapia, R. A. and Thompson, J. R. (1978). Nonparametric Probability Density Estimation, Baltimore: Johns Hopkins University Press.

Tukey, J. W. & McLaughlin, D. H. (1963), Less vulnerable confidence and significance procedures for location based on a single sample: Trimming/Winsorization," Indian Journal of Statistics, 25, 331-351.

Wainer, H., & Thissen, D. (1976). Three steps toward robust regression. Psychometrika, 41, 9-34.

Wegman, E. J. and Carroll, R. J. (1977). A Monte Carlo study of robust estimators of location. Communications in Statistics, A6, 795-812.

Walberg, H. J., Strykowski, B.F., Rovai E., and Hung S.S. (1984). Exceptional performance. Review of Educational Research, 54, 87-112.